Replace this file with prentcsmacro.sty for your meeting, or with entcsmacro.sty for your meeting. Both can be found at the ENTCS Macro Home Page.

Complete Axiomatization for the Total Variation Distance of Markov Chains

Giorgio Bacci[1] Giovanni Bacci Kim G. Larsen Radu Mardare

Department of Computer Science, Aalborg University, Denmark

Abstract

We propose a complete axiomatization for the total variation distance of finite labelled Markov chains. Our axiomatization is given in the form of a quantitative deduction system, a framework recently proposed by Mardare, Panangaden, and Plotkin (LICS 2016) to extend classical equational deduction systems by means of inferences of equality relations $t \equiv_\varepsilon s$ indexed by rationals, expressing that "t is approximately equal to s up to an error ε". Notably, the quantitative equational system is obtained by extending our previous axiomatization (CONCUR 2016) for the probabilistic bisimilarity distance with a distributivity axiom for the prefix operator over the probabilistic choice inspired by Rabinovich's (MFPS 1983).
Finally, we propose a metric extension to the Kleene-style representation theorem for finite labelled Markov chains w.r.t. trace equivalence due to Silva and Sokolova (MFPS 2011).

Keywords: Behavioral Distances, Markov Chains, Axiomatization, Quantitative Deductive Systems.

1 Introduction

In [14], Kleene introduced an algebra of *regular events* to represent behaviors of finite automata. Kleene's approach was essentially equational, but he did not provide a completeness result for his characterization. The first sound and complete axiomatization of regular expressions is due to Salomaa [22], later refined by Kozen [15].

The above programme was applied by Milner [18] to nondeterministic process behaviors, providing a sound and complete axiomatization for bisimilarity of labelled transition systems. Milner's work inspired a number of extensions [5,19,8,23,7], among which it is notable the work of Rabinovich [21] who axiomatized trace equivalence of labelled transition systems. The key observation in Rabinovich's work is that trace equivalence satisfies a distributivity law for the prefix operator over nondeterministic choice. A similar idea was used by Silva and Sokolova [24] for axiomatizing probabilistic trace equivalence of (generative) labelled Markov chains. Their equational characterization extends Stark and Smolka's [25] axiomatization for probabilistic bisimilarity by introducing a distributivity axiom for the prefix operator over probabilistic choice.

[1] Corresponding author: grbacci@cs.aau.dk

The attractiveness towards sound and complete axiomatizations for process behaviors comes from the fact that one can reason about their equivalence in a purely syntactic way by means of classical logical deduction of valid equational statements.

Giacalone, Jou, and Smolka [12,11], however, observed that for reasoning about the behavior of probabilistic systems (more generally, all types of quantitative systems) a notion of distance is preferable to that of equivalence, since the latter is not robust w.r.t. small variations of numerical values. This motivated the development of metric-based semantics for probabilistic systems, initiated by Desharnais et al. [9] on labelled Markov chains and greatly developed and explored by van Breugel, Worrell and others [27,28,3,2]. It consists in proposing a pseudometric which measures the dissimilarities between quantitative behaviors.

Recently, in [4], we provided a sound and complete axiomatization for the probabilistic bisimilarity distance of Desharnais et al. [9] on labelled Markov chains. The axiomatization was given in the form of a *quantitative deduction system*, a concept introduced by Mardare, Panangaden, and Plotkin [17] with the purpose of developing a general research programme for a quantitative algebraic theory of effects [20]. The key idea is to use "quantitative equations" of the form $t \equiv_\varepsilon s$ indexed by positive rationals to be interpreted as "t is approximately equal to s up to an error ε". The main result in [17] is that completeness for a quantitative theory always holds on the freely-generated algebra of terms equipped with a metric that is freely-induced by the axioms.

In this paper, we present a sound and complete axiomatization for the *total variation distance* on labelled Markov chains, *i.e.* a natural metric-lifting of probabilistic trace equivalence. Interestingly, completeness is obtained by extending our axiomatization for the probabilistic bisimilarity distance [4] with the distributivity axiom for the prefix operator over probabilistic choice introduced in [24]. This shows that Rabinovich's idea applies also in a quantitative equational scenario.

The similarities with Rabinovich's work, however, are limited only to the style of the axioms. For our proof of completeness we had to face a number of problems that do not occur in the classical equational case [21,24]. One of the difficulties was that the recursion operator does not satisfy non-expansiveness w.r.t. the total variation distance (*cf.* Example 5.10). This prevented us from directly apply the general completeness result of [17], which specifically requires non-expansiveness for all algebraic operators. This issue was overcome by using a convergence result from [2], that characterizes the total variation distance as the point-wise limit of a convergent net of certain multi-steps probabilistic bisimilarity distances, that approach the total variation distance by extending the probabilistic observations from single labels to *words* on labels of increasing length. Then, by generalizing a proof technique used in [4], we were able to prove completeness for each of these bisimilarity-like distances, and we extended it to the total variation distance (hence, to their point-wise limit) by means of the Archimedean axiom.

As a final result we give a metric extension to the Kleene's representation theorem of Silva and Sokolova [24] (see also [23,6]) for finite labelled Markov chains up-to trace equivalence. Specifically, we show that process expressions, equipped with the pseudometric freely-induced by the axioms, are in *isometric* correspondence with finite labelled Markov chains metrised by the total variation distance.

2 Preliminaries and Notation

For $R \subseteq M \times M$ an equivalence relation, $M/_R$ is its quotient set. For M, N sets, $M \uplus N$ is the disjoint union, $[M \to N]$ (or N^M) the set of functions from M to N.

A *discrete sub-probability* on M is a function $\mu \colon M \to [0,1]$, such that $\mu(M) \leq 1$, where, for $E \subseteq M$, $\mu(E) = \sum_{m \in E} \mu(m)$; it is a *probability distribution* if $\mu(M) = 1$. The support of μ is the set $supp(\mu) = \{m \in M \mid \mu(m) > 0\}$. We denote by $\Delta(M)$ and $\mathcal{D}(M)$ the set of discrete probability and finitely-supported sub-probability distributions on M, respectively.

A 1-bounded *pseudometric* on M is a function $d \colon M \times M \to [0,1]$ such that, for any $m, n, u \in M$, $d(m,m) = 0$, $d(m,n) = d(n,m)$ and $d(m,n) + d(n,u) \geq d(m,u)$; d is a *metric* if, in addition, $d(m,m) = 0$ implies $m = n$. The pair (M, d) is called *(pseudo)metric space*. Recall that, 1-bounded pseudometrics on M ordered point-wise by $d \sqsubseteq d'$ iff $d(m,n) \leq d'(m,n)$, for all $m, n \in M$, form a complete lattice (we denote by $\mathbf{0}$ and $\mathbf{1}$ the bottom and top of 1-bounded pseudometrics, respectively).

3 Quantitative Equational Theories

We recall the main definitions and results from [17]. Let Σ be a signature of function symbols $f \colon n \in \Sigma$ of arity $n \in \mathbb{N}$. Fix a countable set of *metavariables* X, ranged over by $x, y, z, \ldots \in X$. Let $\mathbb{T}(\Sigma, X)$ denote the set of Σ-terms freely generated over X; terms will be ranged over by $t, s, u, \ldots$ A *substitution of type* Σ is a function $\sigma \colon X \to \mathbb{T}(\Sigma, X)$ that is homomorphically extended to terms as $\sigma(f(t_1, \ldots, t_n)) = f(\sigma(t_1), \ldots, \sigma(t_n))$; by $\mathcal{S}(\Sigma)$ we denote the set of substitutions of type Σ.

A *quantitative equation of type* Σ is an expression of the form $t \equiv_\varepsilon s$, where $t, s \in \mathbb{T}(\Sigma, X)$ and $\varepsilon \in \mathbb{Q}_+$. We denote by $\mathcal{E}(\Sigma)$ the set of quantitative equations of type Σ; subsets of $\mathcal{E}(\Sigma)$ will be ranged over by $\Gamma, \Theta, \Pi, \ldots$

Let $\vdash \subseteq 2^{\mathcal{E}(\Sigma)} \times \mathcal{E}(\Sigma)$ be a relation from the powerset of $\mathcal{E}(\Sigma)$ to $\mathcal{E}(\Sigma)$. We write $\Gamma \vdash t \equiv_\varepsilon s$ if $(\Gamma, t \equiv_\varepsilon s) \in \vdash$; by $\vdash t \equiv_\varepsilon s$ we denote $\emptyset \vdash t \equiv_\varepsilon s$, and by $\Gamma \vdash \Theta$ we mean that $\Gamma \vdash t \equiv_\varepsilon s$, for all $t \equiv_\varepsilon s \in \Theta$. The relation $\vdash$ is called *quantitative deduction system of type* Σ if it satisfies the following axioms and rules

$$
\begin{aligned}
&(\mathsf{Refl}) && \vdash t \equiv_0 t \,, \\
&(\mathsf{Symm}) && \{t \equiv_\varepsilon s\} \vdash s \equiv_\varepsilon t \,, \\
&(\mathsf{Triang}) && \{t \equiv_\varepsilon u, u \equiv_{\varepsilon'} s\} \vdash t \equiv_{\varepsilon+\varepsilon'} s \,, \\
&(\mathsf{Max}) && \{t \equiv_\varepsilon s\} \vdash t \equiv_{\varepsilon+\varepsilon'} s \,, \text{ for all } \varepsilon' > 0 \,, \\
&(\mathsf{Arch}) && \{t \equiv_{\varepsilon'} s \mid \varepsilon' > \varepsilon\} \vdash t \equiv_\varepsilon s \,, \\
&(\mathsf{NExp}) && \{t_1 =_\varepsilon s_1, \ldots, t_n =_\varepsilon s_n\} \vdash f(t_1, \ldots, t_n) \equiv_\varepsilon f(s_1, \ldots, s_n) \,, \text{ for all } f \colon n \in \Sigma \,, \\
&(\mathsf{Subst}) && \text{If } \Gamma \vdash t \equiv_\varepsilon s, \text{ then } \sigma(\Gamma) \vdash \sigma(t) \equiv_\varepsilon \sigma(t), \text{ for all } \sigma \in \mathcal{S}(\Sigma) \,, \\
&(\mathsf{Cut}) && \text{If } \Gamma \vdash \Theta \text{ and } \Theta \vdash t \equiv_\varepsilon s, \text{ then } \Gamma \vdash t \equiv_\varepsilon s \,, \\
&(\mathsf{Assum}) && \text{If } t \equiv_\varepsilon s \in \Gamma, \text{ then } \Gamma \vdash t \equiv_\varepsilon s \,.
\end{aligned}
$$

where $\sigma(\Gamma) = \{\sigma(t) \equiv_\varepsilon \sigma(s) \mid t \equiv_\varepsilon s \in \Gamma\}$.

The rules (Subst), (Cut), (Assum) are those of classical logical deduction. The axioms (Refl), (Symm), (Triang) are reflexivity, symmetry, and triangular inequality for a pseudometric, respectively; (Max) is inclusion of neighborhoods of increasing diameter; (Arch) is the Archimedean law; and (NExp) is non-expansiveness of $f \in \Sigma$.

A *quantitative equational theory* is a set $\mathcal{U}$ of universally quantified *quantitative inferences*, (*i.e.*, expressions of the form $\{t_1 \equiv_{\varepsilon_1} s_1, \dots, t_n \equiv_{\varepsilon_n} s_n\} \vdash t \equiv_\varepsilon s$, with a finite set of hypotheses) closed under $\vdash$-deducibility. A set $\mathcal{A}$ of quantitative inferences is said to axiomatize $\mathcal{U}$, if $\mathcal{U}$ is the smallest quantitative equational theory containing $\mathcal{A}$. A theory $\mathcal{U}$ is called *inconsistent* if $\vdash x \equiv_0 y \in \mathcal{U}$, for distinct metavariables $x, y \in X$, it is called *consistent* otherwise[2].

The models of quantitative equational theories are *quantitative algebras*.

Definition 3.1 A *quantitative* Σ-*algebra* is a tuple $\mathcal{A} = (A, \Sigma^{\mathcal{A}}, d^{\mathcal{A}})$, consisting of a pseudometric space $(A, d^{\mathcal{A}})$ and a set $\Sigma^{\mathcal{A}} = \left\{ f^{\mathcal{A}} \colon A^n \to A \mid f \colon n \in \Sigma \right\}$ of *interpretations* for $f \in \Sigma$, required to be non-expansive w.r.t. $d^{\mathcal{A}}$, *i.e.*, for all $1 \le i \le n$ and $a_i, b_i \in A$, $d^{\mathcal{A}}(a_i, b_i) \ge d^{\mathcal{A}}(f^{\mathcal{A}}(a_1, \dots, a_n), f^{\mathcal{A}}(b_1, \dots, b_n))$.

A quantitative algebra $\mathcal{A} = (A, \Sigma^{\mathcal{A}}, d^{\mathcal{A}})$ *satisfies* the inference $\Gamma \vdash t \equiv_\varepsilon s$, written $\Gamma \models_{\mathcal{A}} t \equiv_\varepsilon s$, if for any assignment of the meta-variables $\iota \colon X \to A$,

$$\big(\text{for all } t' \equiv_{\varepsilon'} s' \in \Gamma,\, d^{\mathcal{A}}(\iota(t'), \iota(s')) \le \varepsilon'\big) \qquad \text{implies} \qquad d^{\mathcal{A}}(\iota(t), \iota(s)) \le \varepsilon\,,$$

where, for a term $t \in \mathbb{T}(\Sigma, X)$, $\iota(t)$ denotes the homomorphic interpretation of t in $\mathcal{A}$. A quantitative algebra $\mathcal{A}$ is said to *satisfy* (or is a *model* for) the quantitative theory $\mathcal{U}$, if whenever $\Gamma \vdash t \equiv_\varepsilon s \in \mathcal{U}$, then $\Gamma \models_{\mathcal{A}} t \equiv_\varepsilon s$.

In [17] it is shown that any quantitative theory $\mathcal{U}$ has a universal model $\mathcal{T}_{\mathcal{U}}$ (the freely generated $\vdash$-model) satisfying exactly those quantitative equations belonging to $\mathcal{U}$. Moreover, [17, Theorem 5.2] proves a completeness theorem for quantitative equational theories $\mathcal{U}$, stating that a quantitative inference is satisfied by all the algebras satisfying $\mathcal{U}$ iff it belongs to $\mathcal{U}$.

4 The Algebra of Probabilistic Behaviors

In this section we present the algebra of open Markov chains from [4]. Open Markov chains extend the familiar notion of discrete-time generative Markov chain with "open" states taken from a fixed countable set $\mathcal{X}$ of names ranged over by $X, Y, Z, \dots \in \mathcal{X}$. Names indicate states at which the behavior of the Markov chain can be extended by substitution of another Markov chain, in a way which will be made precise later.

In what follows we fix a countable set $\mathcal{L}$ of labels, ranged over by $a, b, c, \dots$

Definition 4.1 An *open Markov chain* $\mathcal{M} = (M, \tau)$ consists of a set M of *states* and a *transition probability function* $\tau \colon M \to \mathcal{D}((\mathcal{L} \times M) \uplus \mathcal{X})$.

Intuitively, if $\mathcal{M}$ is in a state $m \in M$, then it emits $a \in \mathcal{L}$ and moves to $n \in M$ with probability $\tau(m)(a, n)$, or it moves to a name $X \in \mathcal{X}$ without emitting any label with probability $\tau(m)(X)$. A state $m \in M$ is *terminating* if $\tau(m)((\mathcal{L} \times M) \uplus \mathcal{X}) = 0$. A name $X \in \mathcal{X}$ is *unguarded* in the state $m \in M$, if $\tau(m)(X) > 0$.

[2] Note that for an inconsistent theory $\mathcal{U}$, by Subst, we have $\vdash t \equiv_0 s \in \mathcal{U}$, for all $t, s \in \mathbb{T}(\Sigma, X)$.

A *pointed open Markov chain* $(\mathcal{M}, m) \in \mathbf{OMC}$ is an open Markov chain $\mathcal{M}$ with distinguished *initial* state m. Hereafter, we use $\mathcal{M} = (M, \tau)$, $\mathcal{N} = (N, \theta)$ to denote generic open Markov chains, and $(\mathcal{M}, m)$, $(\mathcal{N}, n)$ for pointed open Markov chains.

Definition 4.2 An equivalence relation $R \subseteq M \times M$ is a *probabilistic bisimulation* on $\mathcal{M}$ if whenever $m \, R \, m'$, then, for all $a \in \mathcal{L}$, $X \in \mathcal{X}$ and $C \in M/_R$,

(i) $\tau(m)(X) = \tau(m')(X)$,

(ii) $\tau(m)(\{a\} \times C) = \tau(m')(\{a\} \times C)$.

Two states $m, m' \in M$ are *probabilistic bisimilar* w.r.t. $\mathcal{M}$, written $m \sim^{\mathcal{M}} m'$, if there exists a probabilistic bisimulation relation on $\mathcal{M}$ relating them.

The above definition is a straightforward adaptation of Larsen and Skou's probabilistic bisimulation [16]. We say that $(\mathcal{M}, m), (\mathcal{N}, n) \in \mathbf{OMC}$ are bisimilar, written $(\mathcal{M}, m) \sim (\mathcal{N}, n)$, if m and n are bisimilar w.r.t. the disjoint union of $\mathcal{M}$ and $\mathcal{N}$ (denoted by $\mathcal{M} \oplus \mathcal{N}$) defined as expected. Note that $\sim$ is an equivalence relation.

Next we turn to a simple algebra of pointed open Markov chains. The signature of algebraic operator symbols is defined as follows.

$$
\begin{aligned}
\Sigma = \{&X \colon 0 \mid X \in \mathcal{X}\} \cup && \text{(NAMES)} \\
&\{a.(\cdot) \colon 1 \mid a \in \mathcal{L}\} \cup && \text{(PREFIX)} \\
&\{+_e \colon 2 \mid e \in [0,1]\} \cup && \text{(PROBABILISTIC CHOICE)} \\
&\{\mathsf{rec}\, X \colon 1 \mid X \in \mathcal{X}\} \,. && \text{(RECURSION)}
\end{aligned}
$$

It consists of a constant X for each name in $\mathcal{X}$; a prefix $a.\cdot$ and a recursion $\mathsf{rec}\, X$ unary operators, for each $a \in \mathcal{L}$ and $X \in \mathcal{X}$; and a probabilistic choice $+_e$ binary operator for each $e \in [0, 1]$. For $t \in \mathbb{T}(\Sigma, M)$, $fn(t)$ denotes the set of free names in t, where the notions of *free* and *bound name* are defined in the standard way, with $\mathsf{rec}\, X$ acting as a binding construct. A term is *closed* if it does not contain any free variable. Throughout the paper we consider two terms as syntactically identical if they are identical up to renaming of their bound names (α-equivalence). For $t, s_1, \dots, s_n \in \mathbb{T}(\Sigma, M)$ and an n-vector $\overline{X} = (X_1, \dots, X_n)$ of distinct names, $t[\overline{s}/\overline{X}]$ denotes the simultaneous *capture avoiding substitution* of X_i in t with s_i, for $i = 1, \dots, n$. A name X is *guarded* [3] in a term t if every free occurrence of X in t occurs within a context the following forms: $a.[\cdot]$, $s +_1 [\cdot]$, or $[\cdot] +_0 s$.

Since from now on we will only refer to terms constructed over the signature Σ, we will simply write $\mathbb{T}(M)$ and $\mathbb{T}$, in place of $\mathbb{T}(\Sigma, M)$ and $\mathbb{T}(\Sigma, \emptyset)$, respectively.

To give the interpretation of the operators in Σ, we define an operator $\mathbb{U}$ on open Markov chains, taking $\mathcal{M}$ to the open Markov chain $\mathbb{U}(\mathcal{M}) = (\mathbb{T}(M), \mu_{\mathcal{M}})$, where the transition probability function $\mu_{\mathcal{M}}$ is defined as the least solution (over the complete partial order of the set of functions mapping elements in $\mathbb{T}(M)$ to a $[0, 1]$-valued function from $(\mathcal{L} \times \mathbb{T}(M)) \uplus \mathcal{X}$, ordered point-wise) of the equation

$$\mu_{\mathcal{M}} = \mathcal{P}_{\mathcal{M}}(\mu_{\mathcal{M}}) \,.$$

[3] This notion, coincides with the one in [25], though our definition may seem more involved due to the fact that we allow the probabilistic choice operators $+_e$ with e ranging in the closed interval $[0, 1]$.

The functional operator $\mathcal{P}_{\mathcal{M}}$ is defined by structural induction on $\mathbb{T}(M)$, for arbitrary functions $\theta\colon \mathbb{T}(M) \to [(\mathcal{L} \times \mathbb{T}(M)) \uplus \mathcal{X} \to [0,1]]$, as follows:

$$\mathcal{P}_{\mathcal{M}}(\theta)(m) = \tau(m) \qquad \mathcal{P}_{\mathcal{M}}(\theta)(a.t) = \mathbb{1}_{\{(a,t)\}}$$
$$\mathcal{P}_{\mathcal{M}}(\theta)(X) = \mathbb{1}_{\{X\}} \qquad \mathcal{P}_{\mathcal{M}}(\theta)(t +_e s) = e\theta(t) + (1-e)\theta(s)$$
$$\mathcal{P}_{\mathcal{M}}(\theta)(\mathsf{rec}\, X.t) = \theta(t[\mathsf{rec}\, X.t/X])\,,$$

where $\mathbb{1}_E$ denotes the characteristic function of the set E.

Notice that requiring $\mu_{\mathcal{M}}$ to be the *least* solution is essential for its well definition as a transition probability function. As a consequence of this definition, for all $X \in \mathcal{X}$, $\mathsf{rec}\, X.X$ is a terminating state in $\mathbb{U}(\mathcal{M})$, i.e., $\mu_{\mathcal{M}}(\mathsf{rec}\, X.X)((\mathcal{L} \times \mathbb{T}(M)) \uplus \mathcal{X}) = 0$.

Remark 4.3 The definition of $\mu_{\mathcal{M}}$ corresponds essentially to the operational semantics given by Stark and Smolka in [25]. The only differences are that the above is given over generic terms in $\mathbb{T}(M)$ and our formulation is simpler because it avoids the construction of a labelled transition system. For more details see [1,25].

Definition 4.4 Let $\mathcal{M}_\emptyset = (\emptyset, \tau_\emptyset)$ be the open Markov chain with $\tau_\emptyset$ the empty transition function. The *universal open Markov chain* is $\mathbb{U}(\mathcal{M}_\emptyset)$.

The reason why it is called universal will be clarified soon. As for now just note that $\mathbb{U}(\mathcal{M}_\emptyset)$ has $\mathbb{T}$ as the set of states and transition probability function equal to the one defined in [25]. To ease the notation we will denote $\mathbb{U}(\mathcal{M}_\emptyset)$ as $\mathbb{U} = (\mathbb{T}, \mu_{\mathbb{T}})$.

The *algebra of open pointed Markov chains* is defined by $(\mathbf{OMC}, \Sigma^{\mathsf{omc}})$, where the interpretations $f^{\mathsf{omc}}\colon \mathbf{OMC}^n \to \mathbf{OMC} \in \Sigma^{\mathsf{omc}}$ for the symbols $f\colon n \in \Sigma$ is given, for arbitrary $(\mathcal{M}, m), (\mathcal{N}, n) \in \mathbf{OMC}$ as

$$X^{\mathsf{omc}} = (\mathbb{U}, X)\,, \qquad (\mathcal{M}, m) +_e^{\mathsf{omc}} (\mathcal{N}, n) = (\mathbb{U}(\mathcal{M} \oplus \mathcal{N}), m +_e n)\,,$$
$$(a.(\mathcal{M}, m))^{\mathsf{omc}} = (\mathbb{U}(\mathcal{M}), a.m)\,, \qquad (\mathsf{rec}\, X.(\mathcal{M}, m))^{\mathsf{omc}} = (\mathbb{U}(\mathcal{M}^*_{X,m}), \mathsf{rec}\, X.m)\,,$$

where, for $\mathcal{M} = (M, \tau)$, $\mathcal{M}^*_{X,m}$ denotes the open Markov chain $(M, \tau^*_{X,m})$ with transition function defined, for all $m' \in M$ and $E \subseteq (\mathcal{L} \times M) \uplus \mathcal{X}$, as

$$\tau^*_{X,m}(m')(E) = \tau(m')(X)\tau(m)(E \setminus \{X\}) + \tau(m')(X^c)\tau(m')(E \setminus \{X\})\,.$$

where $X^c = ((\mathcal{L} \times M) \uplus \mathcal{X}) \setminus \{X\}$. Intuitively, $\tau^*_{X,m}$ removes the name $X \in \mathcal{X}$ from the support of $\tau(m')$ replacing it with the probabilistic behavior of m.

The semantics of terms is given via the Σ-homomorphism $[\![\cdot]\!]\colon \mathbb{T} \to \mathbf{OMC}$, defined by induction on terms as follows

$$[\![X]\!] = X^{\mathsf{omc}}\,, \qquad [\![t +_e s]\!] = [\![t]\!] +_e^{\mathsf{omc}} [\![s]\!]\,,$$
$$[\![a.t]\!] = (a.[\![t]\!])^{\mathsf{omc}}\,, \qquad [\![\mathsf{rec}\, X.t]\!] = (\mathsf{rec}\, X.[\![t]\!])^{\mathsf{omc}}\,. \qquad \text{(SEMANTICS)}$$

Figure 3 shows how terms are interpreted to pointed open Markov chains.

The next result states that it is equivalent to reason about the equivalence of the behavior of $[\![t]\!]$ and $[\![s]\!]$ by just considering bisimilarity between the corresponding states t and s in the universal open Markov chain $\mathbb{U}$.

Theorem 4.5 (Universality) *For all $t \in \mathbb{T}$, $[\![t]\!] \sim (\mathbb{U}, t)$.*

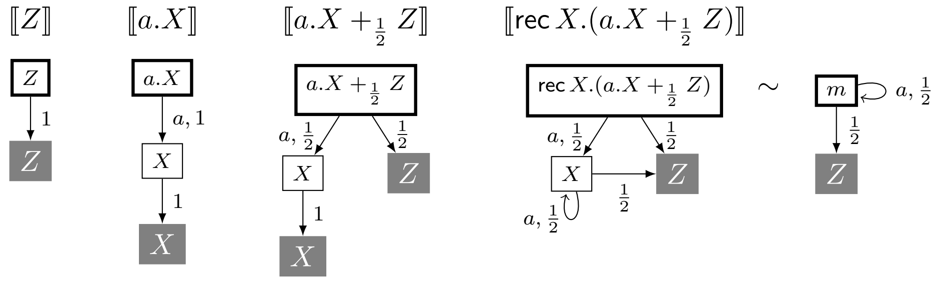


Fig. 1. Compositional step-by-step construction of the semantics of $\mathsf{rec}\,X.(a.X +_{\frac{1}{2}} Z)$.

5 Axiomatization of the Total Variation Distance

We present a quantitative deduction system and we prove its soundness and completeness w.r.t. the total variation distance.

5.1 Probabilistic Trace Equivalence and the Total Variation Distance

To ease the technical presentation of the forthcoming results, it is convenient to interpret sub-probability transition functions $\tau\colon M \to \mathcal{D}((\mathcal{L} \times M) \uplus \mathcal{X})$ fully-probabilistically as $\tau^*\colon M \to \Delta((\mathcal{L} \times M) \uplus \mathcal{X}_\perp)$, where $\mathcal{X}_\perp = \mathcal{X} \uplus \{\perp\}$ and τ^* is the extension of τ such that $\tau^*(m)(\perp) = 1 - \tau(m)((\mathcal{L} \times M) \uplus \mathcal{X})$. These representations are equivalent and do not alter the validity of the previous results.

The type of observable traces that open Markov chains can emit are either finite traces of the form $a_1 \dots a_n \chi \in \mathcal{L}^* \mathcal{X}_\perp$ starting with a sequence of labels $a_1 \dots a_n \in \mathcal{L}^*$ and ending in $\chi \in \mathcal{X}_\perp$ (this indicates the chain ends either in a open name or $\perp$) or infinite sequences of labels $a_1 a_2 a_3 \ldots \in \mathcal{L}^\omega$. We write $\Pi = \mathcal{L}^\omega \cup \mathcal{L}^* \mathcal{X}_\perp$ for the set of observable traces and (Π, Σ_Π) for the measurable space of traces with σ-algebra Σ_Π generated by the set $\mathfrak{C}(\Pi)$ of *cylinders* of the form

$$\mathfrak{C}(w) = \{w\}\Pi \qquad \text{and} \qquad \mathfrak{C}(w\chi) = \{w\chi\}\,, \qquad \text{for } w \in \mathcal{L}^* \text{ and } \chi \in \mathcal{X}_\perp\,.$$

Given a pointed open Markov chain $(\mathcal{M}, m)$, its probability of emitting traces in (Π, Σ_Π) starting from m is given by the distribution $\mathbb{P}_\mathcal{M}(m)$ defined as follows:

Definition 5.1 Let $\mathbb{P}_\mathcal{M}\colon M \to \mathcal{D}(\Pi, \Sigma_\Pi)$ be such that, for all $m \in M$, $\mathbb{P}_\mathcal{M}(m)$ is the *unique* probability measure on (Π, Σ_Π) such that,

$$\begin{aligned} \mathbb{P}_\mathcal{M}(m)(\mathfrak{C}(\chi)) &= \tau^*(m)(\chi)\,, && \text{for all } \chi \in \mathcal{X}_\perp\,, \\ \mathbb{P}_\mathcal{M}(m)(\mathfrak{C}(aw)) &= \textstyle\int \mathbb{P}_\mathcal{M}(\cdot)(\mathfrak{C}(w))\,\mathrm{d}\tau_a(m)\,, && \text{for all } a \in \mathcal{L} \text{ and } w \in \mathcal{L}^* \cup \mathcal{L}^*\mathcal{X}_\perp\,, \end{aligned}$$

where $\tau_a\colon M \to \mathcal{D}(M)$ is given as $\tau_a(m)(n) = \tau(m)(a, n)$, for arbitrary $m, n \in M$.

The existence and unicity of the family of probability measures $(\mathbb{P}_\mathcal{M}(m))_{m \in M}$ satisfying the equations above follows by the Hahn-Kolmogorov extension theorem.

Definition 5.2 Two states $m, m' \in M$ are *probabilistic trace equivalent* w.r.t. $\mathcal{M}$, written $m \approx^\mathcal{M} m'$, if for all cylinders $C \in \mathfrak{C}(\Pi)$, $\mathbb{P}_\mathcal{M}(m)(C) = \mathbb{P}_\mathcal{M}(m')(C)$.

A pseudometric lifting probabilistic trace equivalence is the following.

Definition 5.3 The *total variation distance* $\mathbf{tv}^{\mathcal{M}} \colon M \times M \to [0,1]$ on $\mathcal{M}$ is defined, for arbitrary $m, m' \in M$, as $\mathbf{tv}^{\mathcal{M}}(m,m') = \sup_{E \in \Sigma_\Pi} |\mathbb{P}_{\mathcal{M}}(m)(E) - \mathbb{P}_{\mathcal{M}}(m')(E)|$.

Hereafter, when $\mathcal{M}$ is clear from the context super/subscripts will be omitted.

The proof of completeness in Section 5.5 relies on a result from [2], stating that $\mathbf{tv}$ is the point-wise limit of a net of k-multistep bisimilarity pseudometrics $\mathbf{d}_k$ à la Desharnais et al. [9], whose definition will be made precise later.

Theorem 5.4 ([2]) *Let $\mathbb{K}$ be the poset of positive integers ordered by divisibility. Then, the net $(\mathbf{d}_k)_{k \in \mathbb{K}}$ converges point-wise to $\mathbf{tv}$ and, for all $k \in \mathbb{K}$, $\mathbf{tv} \sqsubseteq \mathbf{d}_k$.*

The intuitive idea of the convergence is that $\mathbf{tv}$ can be approached by stretching the observations from single-step transitions to k-multistep transitions. Here, the pseudometric $\mathbf{d}_k$ measures the k-multistep behavioral similarities of states.

For $k \geq 1$ and $\mathcal{M} = (M, \tau)$ be an open Markov chain, the *k-multistep transition probability function* $\tau_k \colon M \to \Delta((\mathcal{L}^k \times M) \uplus \mathcal{L}^{<k}\mathcal{X}_\perp)$ of $\mathcal{M}$ is defined by induction on k, for arbitrary $m, m' \in M$, $\chi \in \mathcal{X}_\perp$, $a \in \mathcal{L}$, $u \in \mathcal{L}^{<k}$, $w \in \mathcal{L}^k$, as follows

$$\tau_1 = \tau^* \,, \qquad \begin{aligned} \tau_{k+1}(m)(au\chi) &= \textstyle\int \tau_k(\cdot)(u\chi) \, \mathrm{d}\tau_a(m) \,, \\ \tau_{k+1}(m)((aw, m')) &= \textstyle\int \tau_k(\cdot)((w, m')) \, \mathrm{d}\tau_a(m) \,. \end{aligned}$$

The definition of the multistep bisimilarity distances is based on the *Kantorovich (pseudo)metric* between probability distributions over finite (pseudo)metric spaces (A, d), defined for arbitrary $\mu, \nu \in \Delta(A)$ as follows

$$\mathcal{K}(d)(\mu, \nu) = \min \left\{ \textstyle\int d \, \mathrm{d}\omega \mid \omega \in \Omega(\mu, \nu) \right\} \,.$$

where $\Omega(\mu, \nu)$ denotes the set of *couplings* for (μ, ν), *i.e.*, probability distributions $\omega \in \Delta(A \times A)$ such that, for all $E \subseteq A$, $\omega(E \times A) = \mu(E)$ and $\omega(A \times E) = \nu(E)$.

Definition 5.5 For any integer $k \geq 1$, the *probabilistic k-bisimilarity pseudometric* $\mathbf{d}_k^{\mathcal{M}} \colon M \times M \to [0,1]$ on $\mathcal{M}$ is the least fixed-point of the following functional operator on 1-bounded pseudometrics (ordered point-wise),

$$\Psi_k^{\mathcal{M}}(d)(m, m') = \mathcal{K}(\Lambda_k(d))(\tau_k(m), \tau_k(m')) \qquad (k\text{-Kantorovich Operator})$$

where $\Lambda_k(d)$ is the *greatest* 1-bounded pseudometric on $(\mathcal{L}^k \times M) \uplus \mathcal{L}^{<k}\mathcal{X}_\perp$ such that for all $w \in \mathcal{L}^k$ and $m, m' \in M$, $\Lambda_k(d)((w, m), (w, m')) = d(m, m')$.

The well definition of $\mathbf{d}_k$ follows by Knaster-Tarski fixed-point theorem and monotonicity of Ψ_k (which is implied by the monotonicity of $\mathcal{K}$ and Λ_k). Note that $\mathbf{d}_1$ corresponds to the probabilistic bisimilarity distance of Desharnais et al. [9,27].

Definition 5.6 An equivalence relation $R \subseteq M \times M$ is a *k-probabilistic bisimulation* on $\mathcal{M}$ if whenever $m \, R \, m'$, then, for all $w \in \mathcal{L}^k$, $u \in \mathcal{L}^{<k}$, $\chi \in \mathcal{X}_\perp$, and $C \in M/_R$,

(i) $\tau_k(m)(u\chi) = \tau_k(m')(u\chi)$,

(ii) $\tau_k(m)(\{w\} \times C) = \tau_k(m')(\{w\} \times C)$.

Two states $m, m' \in M$ are *k-probabilistic bisimilar* w.r.t. $\mathcal{M}$, written $m \sim_k^{\mathcal{M}} m'$, if there exists a k-probabilistic bisimulation relation on $\mathcal{M}$ relating them.

Lemma 5.7 $\mathbf{d}_k(m,m') = 0$ *iff* $m \sim_k m'$.

For each $k \geq 1$, the k-bisimilarity distance can be alternatively obtained as $\mathbf{d}_k = \bigsqcap_{i \in \omega} \tilde{\Psi}_k^i(\mathbf{1})$, *i.e.*, the ω-limit of the decreasing chain $\mathbf{1} \sqsupseteq \tilde{\Psi}_k(\mathbf{1}) \sqsupseteq \tilde{\Psi}_k^2(\mathbf{1}) \sqsupseteq \dots$, where $\tilde{\Psi}_k$ is the operator

$$\tilde{\Psi}_k(d)(m,m') = \begin{cases} 0 & \text{if } m \sim_k m', \\ \Psi_k(d)(m,m') & \text{otherwise.} \end{cases}$$

Lemma 5.8 $\tilde{\Psi}_k$ *is* ω*-cocontinuous,* i.e.*, for any countable decreasing chain* $d_0 \sqsupseteq d_1 \sqsupseteq d_2 \sqsupseteq \dots$*, it holds* $\bigsqcap_{i \in \omega} \tilde{\Psi}_k(d_i) = \tilde{\Psi}_k(\bigsqcap_{i \in \omega} d_i)$*. Moreover,* $\mathbf{d}_k = \bigsqcap_{i \in \omega} \tilde{\Psi}_k^i(\mathbf{1})$*.*

5.2 A Quantitative Algebra of Open Markov Chains

We turn the algebra of pointed open Markov chains $(\mathbf{OMC}, \Sigma^{\mathsf{omc}})$ given in Section 4 into a *relaxed* quantitative algebra by endowing it with the total variation distance. The term "relaxed" is used to stress the fact that differently from Definition 3.1, we do not require non-expansiveness for the algebraic operators. This relaxation is necessary because the recursion operator does not satisfy non-expansiveness.

We define the total variation distance $\mathbf{tv}^{\mathsf{omc}}$ on $\mathbf{OMC}$ as the total variation distance between the initial states on the disjoint union of the two open Markov chains. Equivalently, one can compute it simply as

$$\mathbf{tv}^{\mathsf{omc}}((\mathcal{M},m),(\mathcal{N},n)) = \sup_{E \in \Sigma_\Pi} |\mathbb{P}_{\mathcal{M}}(m)(E) - \mathbb{P}_{\mathcal{N}}(n)(E)| \,.$$

We define the (relaxed) *quantitative algebra of* $\mathbf{OMC}$ as $(\mathbf{OMC}, \Sigma^{\mathsf{omc}}, \mathbf{tv}^{\mathsf{omc}})$.

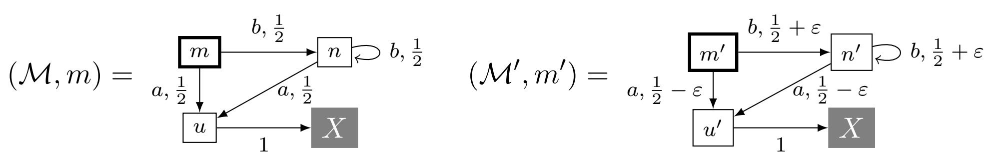
In Example 5.10 we show that the operator $\mathsf{rec}\, X$ fails to be non-expansive w.r.t. the total variation distance. Our proof relies on the following technical lemma.

Lemma 5.9 *For arbitrary* $m, n \in M$*,* $\mathbf{tv}(m,n) < 1$ *iff one of the following holds*

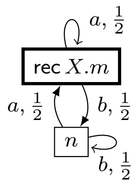
(i) $\mathbb{P}(m)(\mathfrak{C}(w)) > 0$ *and* $\mathbb{P}(n)(\mathfrak{C}(w)) > 0$ *for some* $w \in \mathcal{L}^*\mathcal{X}_\perp$*,*

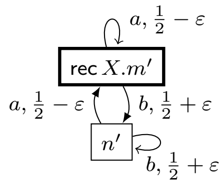
(ii) $\tau_{|w|}(m)((w,m')) > 0$ *and* $\tau_{|w|}(n)((w,n')) > 0$*, for some* $m' \approx n'$ *and* $w \in \mathcal{L}^*$*.*

Example 5.10 (Recursion is not non-expansive!) Let $0 < \varepsilon < \frac{1}{2}$ and consider the two pointed open Markov chains depicted below.



Since $\mathbb{P}(m)(\mathfrak{C}(aX)) = \frac{1}{2}$ and $\mathbb{P}(m')(\mathfrak{C}(aX)) = \frac{1}{2} - \varepsilon > 0$, by Lemma 5.9 (i), $\mathbf{tv}(m,m') < 1$. Consider the application of the operator $(\mathsf{rec}\, X)^{\mathsf{omc}}$ on these two pointed open Markov chains. The resulting chains have the following behavior.

$(\mathsf{rec}\,X.(\mathcal{M}, m))^{\mathsf{omc}} \sim$ 

$(\mathsf{rec}\,X.(\mathcal{M}', m'))^{\mathsf{omc}} \sim$ 

Note that $\mathbb{P}(\mathsf{rec}\,X.m)(\mathfrak{C}(u)) = \mathbb{P}(\mathsf{rec}\,X.m')(\mathfrak{C}(u)) = 0$, for all $u \in \mathcal{L}^*\mathcal{X}_\perp$ (the chains are fully-probabilistic and there are no "open states"). Moreover, by the assumption that $0 < \varepsilon < \frac{1}{2}$, none of the states in $(\mathsf{rec}\,X.(\mathcal{M}, m))^{\mathsf{omc}}$ is trace equivalent with any of the states in $(\mathsf{rec}\,X.(\mathcal{M}', m'))^{\mathsf{omc}}$. Hence, by Lemma 5.9, $\mathbf{tv}((\mathsf{rec}\,X.(\mathcal{M}, m))^{\mathsf{omc}}, (\mathsf{rec}\,X.(\mathcal{M}', m'))^{\mathsf{omc}}) = 1$. This proves that $(\mathsf{rec}\,X)^{\mathsf{omc}}$ fails to be non-expansive w.r.t. the total variation distance.

The following result is a direct consequence of Theorem 4.5 and the fact that bisimilarity implies trace equivalence.

Theorem 5.11 *For all* $t, s \in \mathbb{T}$, $\mathbf{tv}^{\mathsf{omc}}([\![t]\!], [\![s]\!]) = \mathbf{tv}^{\mathbb{U}}(t, s)$.

5.3 A Quantitative Deduction System

Now we present a relaxed quantitative deduction system which will be later shown to be sound and complete w.r.t. the total variation distance. The deduction system we propose is not a quantitative deduction system in the sense of [17], because it does not include the (NExp) axiom of non-expansiveness of the operators (*cf.* Section 3).

The quantitative deduction system $\vdash \subseteq 2^{\mathcal{E}(\Sigma)} \times \mathcal{E}(\Sigma)$ of type Σ that we consider satisfies the axioms (Refl), (Symm), (Triang), (Max), (Arch) and rules (Subst), (Cut) (Assum) from Section 3 and the following additional axioms

$$
\begin{aligned}
&(\mathsf{B1}) && \vdash t +_1 s \equiv_0 t\,,\\
&(\mathsf{B2}) && \vdash t +_e t \equiv_0 t\,,\\
&(\mathsf{SC}) && \vdash t +_e s \equiv_0 s +_{1-e} t\,,\\
&(\mathsf{SA}) && \vdash (t +_e s) +_{e'} u \equiv_0 t +_{ee'} (s +_{\frac{e'-ee'}{1-ee'}} u)\,, \text{ for } e, e' \in [0,1)\,,\\
&(\mathsf{Unfold}) && \vdash \mathsf{rec}\,X.t \equiv_0 t[\mathsf{rec}\,X.t/X]\,,\\
&(\mathsf{Unguard}) && \vdash \mathsf{rec}\,X.(t +_e X) \equiv_0 \mathsf{rec}\,X.t\,,\\
&(\mathsf{Fix}) && \{s \equiv_0 t[s/X]\} \vdash s \equiv_0 \mathsf{rec}\,X.t\,, \text{ for } X \text{ guarded in } t\,,\\
&(\mathsf{Cong}) && \{t \equiv_0 s\} \vdash \mathsf{rec}\,X.t \equiv_0 \mathsf{rec}\,X.s\,,\\
&(\mathsf{DP}) && \vdash a.(t +_e s) \equiv_0 a.t +_e a.s\,,\\
&(\mathsf{Top}) && \vdash t \equiv_1 s\,,\\
&(\mathsf{Pref}) && \{t \equiv_\varepsilon s\} \vdash a.t \equiv_\varepsilon a.s\,,\\
&(\mathsf{IB}) && \{t \equiv_\varepsilon s, t' \equiv_{\varepsilon'} s'\} \vdash t +_e t' \equiv_{\varepsilon''} s +_e s'\,, \text{ for } \varepsilon'' \geq e\varepsilon + (1-e)\varepsilon'\,.
\end{aligned}
$$

Note that the axiom (NExp) is not included in the definition.

(B1), (B2), (SC), (SA) are the axioms of *barycentric algebras* due to Stone [26], used here to axiomatize the convex set of probability distributions. (SC) stands for *skew commutativity* and (SA) for *skew associativity*.

The axioms (Unfold), (Unguard), (Fix), (Cong) are the recursion axioms of Milner [18], used here to axiomatize coinductive behaviors. (Unfold) and (Fix) state that, whenever X is guarded in a term t, $\mathsf{rec}\, X.t$ is *the unique solution* of the recursive equation $s \equiv_0 t[s/X]$. The axiom (Unguard) deals with unguarded recursive behavior, and (Cong) states the congruential properties of the recursion operator.

The axiom (DP) is the distributivity law of the prefix operator over the probabilistic choice of Silva and Sokolova [24].

The last three axioms are the only truly quantitative one. (Top) states that the distance is bounded by 1; (Pref) is the non-expansiveness for the prefix operator; and (IB) is the *interpolative barycentric axiom* introduced in [17] for axiomatizing the Kantorovich distance on finitely-supported probability distributions (*cf.* §10 in [17]).

It is important to remark that the quantitative deduction system given above subsumes the equational systems of Stark and Smolka [25] axiomatizing probabilistic bisimilarity, and of Silva and Sokolova [24] for probabilistic trace equivalence.

5.4 Soundness

In this section we show the soundness of our quantitative deduction system w.r.t. the bisimilarity distance between pointed open Markov chains.

Recall that, by Theorem 5.11, for all $t, s \in \mathbb{T}$, $\models_{\mathbf{OMC}} t \equiv_\varepsilon s$ is equivalent to $\models_{\mathbb{U}} t \equiv_\varepsilon s$. To ease the notation, in the following we simply write $\models t \equiv_\varepsilon s$.

Theorem 5.12 (Soundness) *For any $t, s \in \mathbb{T}$, if $\vdash t \equiv_\varepsilon s$ then $\models t \equiv_\varepsilon s$.*

Proof. The axioms (Refl), (Symm), (Triang), (Max), and (Arch) are sound since $\mathbf{tv}$ is a pseudometric. The soundness of the classical logical deduction rules (Subst), (Cut), and (Assum) is immediate. By $\sim \subseteq \approx$, the axioms (B1), (B2), (SC), and (SA) along with (Unfold) and (Unguard) follow directly by the soundness theorem proven in [25]. The soundness of (Cong), (Fix), (DP) follow by the soundness of [24].

The soundness of (Top) follows from the fact that $\mathbf{tv}$ is 1-bounded. To prove the soundness of (Pref) it suffices to show $\mathbf{tv}(t, s) \geq \mathbf{tv}(a.t, a.s)$.

$$
\begin{aligned}
\mathbf{tv}(a.t, a.s) &= \sup_{E \in \Sigma_\Pi} |\mathbb{P}(a.t)(E) - \mathbb{P}(a.s)(E)| && \text{(def. } \mathbf{tv}) \\
&= \sup_{E \in \Sigma_\Pi} |\mathbb{P}(a.t)(\{a\}E) - \mathbb{P}(a.s)(\{a\}E)| && (\mathbb{P}(a.t)(\mathfrak{C}(a)) = \mathbb{P}(a.s)(\mathfrak{C}(a)) = 1) \\
&= \sup_{E \in \Sigma_\Pi} |\mathbb{P}(t)(E) - \mathbb{P}(s)(E)| = \mathbf{tv}(t, s) && \text{(def. } \mu_{\mathbb{T}},\ \mathcal{P}_{\mathbb{U}},\ \mathbf{tv})
\end{aligned}
$$

The soundness of (IB) follows by $e\,\mathbf{tv}(t, s) + (1 - e)\,\mathbf{tv}(t', s') \geq \mathbf{tv}(t +_e t', s +_e s')$.

$$
\begin{aligned}
\mathbf{tv}(t +_e t', s +_e s') &= \sup_{E \in \Sigma_\Pi} |\mathbb{P}(t +_e t)(E) - \mathbb{P}(s +_e s')(E)| && \text{(def. } \mathbf{tv}) \\
&= \sup_{E \in \Sigma_\Pi} \big|e\big(\mathbb{P}(t)(E) - \mathbb{P}(s)(E)\big) + (1 - e)\big(\mathbb{P}(t')(E) - \mathbb{P}(s')(E)\big)\big| && \text{(def. } \mu_{\mathbb{T}},\ \mathcal{P}_{\mathbb{U}}) \\
&\leq e \sup_{E \in \Sigma_\Pi} |\mathbb{P}(t)(E) - \mathbb{P}(s)(E)| + (1 - e) \sup_{E \in \Sigma_\Pi} |\mathbb{P}(t')(E) - \mathbb{P}(s')(E)| \\
&\leq e\,\mathbf{tv}(t, s) + (1 - e)\,\mathbf{tv}(t', s') && \text{(def. } \mathbf{tv})
\end{aligned}
$$

The above concludes the proof. □

5.5 Completeness

In this section we prove completeness of our quantitative deduction system w.r.t. the total variation distance between pointed open Markov chains.

For the sake of readability we introduce the following notation for *formal sums of terms* (or *convex combinations of terms*). For $n \geq 1$, $t_1, \dots, t_n \in \mathbb{T}$ terms, and $e_1, \dots, e_n \in [0,1]$ positive reals such that $\sum_{i=1}^{n} e_i = 1$, we define

$$\textstyle\sum_{i=1}^{n} e_i \cdot t_i = \begin{cases} t_1 & \text{if } e_1 = 1 \\ t_1 +_{e_1} \left(\sum_{i=2}^{n} \frac{e_i}{1-e_1} \cdot t_i \right) & \text{otherwise}\,. \end{cases}$$

Following the pattern of [18,25], the completeness theorem hinges on a couple of important transformations. The first of these is the Bekič-Scott construction of solutions of simultaneous recursive definitions. This is embodied in the next theorem, which is [18, Theorem 5.7].

Theorem 5.13 (Unique Solution of Equations) *Let $\overline{X} = (X_1, \dots, X_k)$ and $\overline{Y} = (Y_1, \dots, Y_h)$ be distinct names, and $\overline{t} = (t_1, \dots, t_k)$ terms with free names in $(\overline{X}, \overline{Y})$ in which each X_i is guarded. Then there exist terms $\overline{s} = (s_1, \dots, s_k)$ with free names in $\overline{Y}$ such that*

$$\vdash s_i \equiv_0 t_i[\overline{s}/\overline{X}]\,, \qquad\qquad \textit{for all } i \leq k.$$

Moreover, if for some $\overline{u} = (u_1, \dots, u_k)$ with free variables in $\overline{Y}$, $\vdash u_i \equiv_0 t_i[\overline{u}/\overline{X}]$, for all $i \leq k$, then $\vdash s_i \equiv_0 u_i$, for all $i \leq k$.

The second transformation provides a deducible "k-steps normal form" for terms, where k is a parameter counting the number of nested prefixes. This result is embodied in the following theorem (when $k = 1$, this is [25, Theorem 5.9] [4]).

Theorem 5.14 (Eq. Characterization) *For any $k \geq 1$ and any term t, with free names in $\overline{Y}$, there exist terms $t_1, \dots, t_n$ with free names in $\overline{Y}$, s.t. $\vdash t \equiv_0 t_1$ and*

$$\vdash t_i \equiv_0 \textstyle\sum_{j=1}^{h(i)} p_{ij} \cdot s_{ij}\,, \qquad\qquad \textit{for all } i \leq n\,,$$

where the terms s_{ij} are enumerated without repetitions, each s_{ij} is either of the form $a^1_{ij}.\dots.a^k_{ij}.t_{f(i,j)}$, or $b^1_{ij}.\dots.b^{u_{ij}}_{ij}.\,\mathsf{rec}\,X.X$, or $c^1_{ij}.\dots.c^{w_{ij}}_{ij}.Y_{g(i,j)}$, for some $u_{ij}, w_{ij} < k$.

The last lemma relates the proposed deduction system with the Kantorovich distance. So far this is the only transformation embodying the use of the interpolative barycentric axiom (IB) to deduce nontrivial quantitative information on terms.

---

[4] The formulation given here is slightly simpler than the original one in [25], since our deduction system satisfies the axiom (B1), which is not included in the equational deduction system of [25].

Lemma 5.15 *Let d be a 1-bounded pseudometric over $\mathbb{T}$ and $\mu, \nu \in \Delta(\mathbb{T})$ probability measures with supports $supp(\mu) = \{t_1, \dots, t_k\}$ and $supp(\nu) = \{s_1, \dots, s_r\}$. Then*

$$\{t_i \equiv_\varepsilon s_u \mid \varepsilon \geq d(t_i, s_u), 1 \leq i \leq k \text{ and } 1 \leq u \leq r\} \vdash \sum_{i=1}^{k} \mu(t_i) \cdot t_i \equiv_{\varepsilon'} \sum_{u=1}^{r} \nu(s_u) \cdot s_u \,,$$

for all $\varepsilon' \geq \mathcal{K}(d)(\mu, \nu)$.

Now we are ready to prove the main result of this section. The proof of completeness can be sketched as follows. Given $t, s \in \mathbb{T}$ such that $\mathbf{tv}(t, s) \leq \varepsilon$, to prove $\vdash t \equiv_\varepsilon s$, we first show that, for any $k \geq 1$, $\varepsilon_k \geq \mathbf{d}_k(t, s)$ implies $\vdash t \equiv_{\varepsilon_k} s$, then, by Theorem 5.4 and (Arch), we deduce $\vdash t \equiv_\varepsilon s$.

Theorem 5.16 (Completeness) *For any $t, s \in \mathbb{T}$, if $\models t \equiv_\varepsilon s$, then $\vdash t \equiv_\varepsilon s$.*

Proof. Let $t, s \in \mathbb{T}$ and $\varepsilon \in \mathbb{Q}_+$. By Theorem 5.4 and (Arch), to prove that $\mathbf{tv}(t, s) \leq \varepsilon$ implies $\vdash t \equiv_\varepsilon s$, it suffices to show that for any $k \geq 1$, $\varepsilon \geq \mathbf{d}_k(t, s)$ implies $\vdash t \equiv_\varepsilon s$. Let $k \geq 1$. The case $\varepsilon \geq 1$ trivially follows by (Top) and (Max). Let $\varepsilon < 1$. By Theorem 5.14, there exist terms $t_1, \dots, t_m$ and $s_1, \dots, s_r$ with free names in $\overline{X}$ and $\overline{Y}$, respectively, such that $\vdash t \equiv_0 t_1$, $\vdash s \equiv_0 s_1$, and

$$\vdash t_i \equiv_0 \textstyle\sum_{j=1}^{h(i)} p_{ij} \cdot t'_{ij} \,, \qquad \text{for all } i \leq m, \tag{1}$$

$$\vdash s_u \equiv_0 \textstyle\sum_{v=1}^{n(u)} q_{uv} \cdot s'_{uv} \,, \qquad \text{for all } u \leq r, \tag{2}$$

where the terms t'_{ij} (resp. s'_{uv}) are enumerated without repetitions, and t'_{ij} (resp. s'_{uv}) have either the form $a^1_{ij}.\dots.a^k_{ij}.t_{f(i,j)}$ (resp. $b^1_{uv}.\dots.b^k_{uv}.s_{z(u,v)}$), or $a^1_{ij}.\dots.a^{\alpha_{ij}}_{ij}.\,\mathsf{rec}\, Z.Z$ (resp. $b^1_{uv}.\dots.b^{\beta_{uv}}_{uv}.\,\mathsf{rec}\, Z.Z$) , or $a^1_{ij}.\dots.a^{\gamma_{ij}}_{ij}.X_{g(i,j)}$ (resp. $b^1_{uv}.\dots.b^{\delta_{uv}}_{uv}.Y_{w(u,v)}$), for some $\alpha_{ij}, \gamma_{ij} < k$ (resp. $\beta_{uv}, \delta_{uv} < k$). If we can prove that for all $\sigma \in \omega$,

$$\vdash t_i \equiv_\varepsilon s_u \,, \qquad \text{for all } i \leq m,\ u \leq r, \text{ and } \varepsilon \geq \tilde{\Psi}^\sigma_k(\mathbf{1})(t_i, s_u) \,, \tag{3}$$

then, by Lemma 5.8 and (Arch), we get $\vdash t_i \equiv_\varepsilon s_u$, for all $\varepsilon \geq \mathbf{d}_k(t_i, s_u)$. Since $\vdash t \equiv_0 t_1$, $\vdash s \equiv_0 s_1$, by (Triang), we deduce $\vdash t \equiv_\varepsilon s$, for all $\varepsilon \geq \mathbf{d}_k(t, s)$.

In the remainder of the proof we prove (3), by induction on $\sigma \in \omega$.

(Base case: $\sigma = 0$) $\tilde{\Psi}^0_k(\mathbf{1})(t_i, s_u) = \mathbf{1}(t_i, s_u)$. Since $\mathbf{1}(t_i, s_u) = 0$ whenever $t_i = s_u$ and $\mathbf{1}(t_i, s_u) = 1$ if $t_i \neq s_u$, then (3) follows by the axioms (Refl), (Top) and (Max).

(Inductive step: $\sigma > 0$). We consider the cases $t_i \sim_k s_u$ and $t_i \not\sim_k s_u$ separately. Assume $t_i \sim_k s_u$. Since our deduction system includes the one of Silva and Sokolova and $\sim_k \subseteq \approx$, whenever $t_i \sim_k s_u$, by completeness w.r.t. $\approx$ ([24, Theorem 5.11]), we get $\vdash t_i \equiv_0 s_u$. By (Max), $\vdash t_i \equiv_\varepsilon s_u$, for all $\varepsilon \geq \tilde{\Psi}^\sigma_k(\mathbf{1})(t_i, s_u) = 0$. Assume $t_i \not\sim_k s_u$ and let t'_{ij} and s'_{uv} be the terms occurring in the formal sums of (1), (2), respectively. Then, by (1), (2), and definition of $(\mu_\mathbb{T})_k$, for $x, y \in (\mathcal{L}^k \times \mathbb{T}) \uplus \mathcal{L}^{<k}\mathcal{X}_\bot$,

$$(\mu_\mathbb{T})_k(t_i)(x) = \begin{cases} p_{ij} & \text{if } \xi(x) = t'_{ij} \\ 0 & \text{otherwise}\,, \end{cases} \qquad (\mu_\mathbb{T})_k(s_u)(y) = \begin{cases} q_{uv} & \text{if } \xi(y) = s'_{uv} \\ 0 & \text{otherwise}\,. \end{cases}$$

where ξ is the mapping s.t. for all $t \in \mathbb{T}$, $a_1, \dots, a_k \in \mathcal{L}$, $k' < k$, and $X \in \mathcal{X}$, $\xi((a_1 \dots a_k, t)) = a_1.\dots.a_k.t$, $\xi(a_1 \dots a_{k'} X) = a_1.\dots.a_{k'}.X$, and $\xi(a_1 \dots a_{k'} \bot) =$

$a_1 \ldots a_{k'}. \mathsf{rec}\, Z.Z$. If we can prove that, for all $i \le m$, $j \le h(i)$, $u \le r$, and $v \le n(u)$,

$$\vdash t'_{ij} \equiv_\varepsilon s'_{uv}\,, \qquad\qquad \text{for all } \varepsilon \ge \Lambda_k(\tilde{\Psi}_k^{\sigma-1}(\mathbf{1}))(\xi(t'_{ij}), \xi(s'_{uv}))\,, \tag{4}$$

then, by Lemma 5.15, for all $\varepsilon \ge \mathcal{K}(\Lambda_k(\tilde{\Psi}_k^{\sigma-1}(\mathbf{1})))((\mu_{\mathbb{T}})_k(t_i), (\mu_{\mathbb{T}})_k(s_u))$ we can deduce $\vdash t_i \equiv_\varepsilon s_u$. By this and definitions of $\tilde{\Psi}_k^\sigma$ and Ψ_k, since we assumed $t_i \not\sim_k s_u$, we conclude (3).

Next we prove (4). The only interesting case is when $t'_{i,j} = a_1 \ldots a_k . t_{f(i,j)}$ and $s'_{u,v} = a_1 \ldots a_k . s_{z(u,v)}$ —the others follow by (Refl), if $t'_{i,j} = s'_{u,v}$, (Top) otherwise, and then (Max). By definitions of ξ and Λ_k, we have $\Lambda_k(\tilde{\Psi}^{\sigma-1}(\mathbf{1}))(\xi(t'_{ij}), \xi(s'_{uv})) = \tilde{\Psi}^{\sigma-1}(\mathbf{1})(t_{f(i,j)}, s_{z(u,v)})$. By inductive hypothesis on $\sigma - 1$, we have $\vdash t_{f(i,j)} \equiv_\varepsilon s_{z(u,v)}$, for all $\varepsilon \ge \tilde{\Psi}^{\sigma-1}(\mathbf{1})(t_{f(i,j)}, s_{z(u,v)})$, and by repeatedly applying (Pref) we get (4). □

6 A Quantitative Kleene's Theorem

In this section we provide a metric extension to the quantitative Kleene's representation theorem of Silva and Sokolova [24] (see also [23,6]). Specifically, we have that any (finite) pointed open Markov chains can be represented up to bisimilarity as a Σ-term and, *vice versa*, for any Σ-term t there exist a (finite) pointed open Markov chain bisimilar to $[\![t]\!]$. Moreover, by endowing the set of Σ-terms with the pseudometric freely-generated by the quantitative deduction system $\vdash$ presented in Section 5.3 (in a way which will be made precise later in Definition 6.1) we get that the correspondence stated above is *metric invariant*.

Next we recall from [17] the definition of initial quantitative model for a quantitative deduction system.

Definition 6.1 The *initial $\vdash$-model* is the quantitative algebra $(\mathbb{T}, \Sigma, d_{\mathbb{T}})$, where $(\mathbb{T}, \Sigma)$ is the initial algebra of Σ-terms and $d_{\mathbb{T}} \colon \mathbb{T} \times \mathbb{T} \to [0,1]$ is the 1-bounded pseudometric defined, for arbitrary terms $t, s \in \mathbb{T}$, as $d_{\mathbb{T}}(t,s) = \inf \{\varepsilon \mid \vdash t \equiv_\varepsilon s\}$.

By (Refl), (Symm), (Triang), (Top) it is easy to prove that $d_{\mathbb{T}}$ is a well-defined 1-bounded pseudometric. Moreover, $(\mathbb{T}, \Sigma, d_{\mathbb{T}})$ is a sound model for $\vdash$.

Next we show that there is a strong correspondence between the initial $\vdash$-model and the quantitative algebra of finite pointed open Markov chains.

Theorem 6.2 (Quantitative Kleene's Theorem)

(i) For any pair $(\mathcal{M}, m), (\mathcal{N}, n)$ of finite open Markov chains, there exist $t, s \in \mathbb{T}$ such that $[\![t]\!] \sim (\mathcal{M}, m)$, $[\![s]\!] \sim (\mathcal{N}, n)$, and $\mathbf{tv}((\mathcal{M}, m), (\mathcal{N}, n)) = d_{\mathbb{T}}(t,s)$;

(ii) For any $t, s \in \mathbb{T}$, there exist finite pointed open Markov chains $(\mathcal{M}, m), (\mathcal{N}, n)$, such that $[\![t]\!] \sim (\mathcal{M}, m)$, $[\![s]\!] \sim (\mathcal{N}, n)$, and $\mathbf{tv}((\mathcal{M}, m), (\mathcal{N}, n)) = d_{\mathbb{T}}(t,s)$.

7 Conclusions and Future Work

In this paper we proposed a sound and complete axiomatization for the total variation distance of Markov chains. The proposed axiomatic system comes as a natural generalization of the one in [24] for probabilistic trace equivalence.

Similarly to [4], where we provided a sound and complete axiomatization for the bisimilarity distance of Desharnais et al., also this case recursion was not sound w.r.t. the non-expansiveness axiom (NExp). Still we were able to prove completeness for the axiomatization. This further result entails the possibility of generalizing the original quantitative framework of Mardare, Panangaden, and Plotkin [17], maybe allowing for algebraic operators that are simply required to be continuous.

Another appealing direction of future work is to apply our results on quantitative systems described as coalgebras in a way similar to one proposed in [6,23,7]. By pursuing this direction we would be able to obtain metric axiomatizations for a wide variety of quantitative systems, including non-generative probabilistic models, weighted transition systems, Segala's systems, stratified systems, Pnueli-Zuck systems, etc.

References

[1] Luca Aceto, Zoltán Ésik, and Anna Ingólfsdóttir. Equational axioms for probabilistic bisimilarity. In *AMAST 2002*, Lecture Notes in Computer Science, pages 239–253, 2002.

[2] Giorgio Bacci, Giovanni Bacci, Kim G. Larsen, and Radu Mardare. Converging from Branching to Linear Metrics on Markov Chains. In *ICTAC*, volume 9399 of *Lecture Notes in Computer Science*, pages 349–367. Springer, 2015.

[3] Giorgio Bacci, Giovanni Bacci, Kim G. Larsen, and Radu Mardare. On the Total Variation Distance of Semi-Markov Chains. In *FoSSaCS*, volume 9034 of *Lecture Notes in Computer Science*, pages 185–199, 2015.

[4] Giorgio Bacci, Giovanni Bacci, Kim G. Larsen, and Radu Mardare. Complete Axiomatization for the Bisimilarity Distance on Markov Chains. In *CONCUR*, volume 59 of *LIPIcs*, pages 21:1–21:14. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2016.

[5] Emanuele Bandini and Roberto Segala. Axiomatizations for probabilistic bisimulation. In *ICALP*, Lecture Notes in Computer Science, pages 370–381, 2001.

[6] Filippo Bonchi, Marcello M. Bonsangue, Jan J. M. M. Rutten, and Alexandra Silva. Deriving syntax and axioms for quantitative regular behaviours. In *CONCUR*, volume 5710 of *Lecture Notes in Computer Science*, pages 146–162. Springer, 2009.

[7] Marcello Bonsangue, Stefan Milius, and Alexandra Silva. Sound and Complete Axiomatizations of Coalgebraic Language Equivalence. *ACM Transactions on Computational Logic*, 14(1):7, 2013.

[8] Yuxin Deng and Catuscia Palamidessi. Axiomatizations for probabilistic finite-state behaviors. *Theor. Comput. Sci.*, 373(1-2):92–114, 2007.

[9] Josee Desharnais, Vineet Gupta, Radha Jagadeesan, and Prakash Panangaden. Metrics for labelled Markov processes. *Theoretical Computer Science*, 318(3):323–354, 2004.

[10] Norm Ferns, Prakash Panangaden, and Doina Precup. Metrics for finite Markov Decision Processes. In *UAI*, pages 162–169. AUAI Press, 2004.

[11] Alessandro Giacalone, Chi chang Jou, and Scott A. Smolka. Algebraic reasoning for probabilistic concurrent systems. In *Proc. IFIP TC2 Working Conference on Programming Concepts and Methods*, pages 443–458. North-Holland, 1990.

[12] Chi-Chang Jou and Scott A. Smolka. Equivalences, congruences, and complete axiomatizations for probabilistic processes. In *CONCUR*, volume 458 of *Lecture Notes in Computer Science*, pages 367–383. Springer, 1990.

[13] Klaus Keimel and Gordon D. Plotkin. Mixed powerdomains for probability and nondeterminism. *Logical Methods in Computer Science*, Volume 13, Issue 1, 2017.

[14] Stephen Kleene. Representation of events in nerve nets and finite automata. *Automata Studies*, 34:3–42, 1956.

[15] Dexter Kozen. A completeness theorem for kleene algebras and the algebra of regular events. In *LICS*, pages 214–225. IEEE Computer Society, 1991.

[16] Kim Guldstrand Larsen and Arne Skou. Bisimulation Through Probabilistic Testing. In *POPL*, pages 344–352, 1989.

[17] Radu Mardare, Prakash Panangaden, and Gordon D. Plotkin. Quantitative Algebraic Reasoning. In *LICS*, pages 700–709. ACM, 2016.

[18] Robin Milner. A complete inference system for a class of regular behaviours. *J. Comput. Syst. Sci.*, 28(3):439–466, 1984.

[19] Michael Mislove, Joël Ouaknine, and James Worrell. Axioms for probability and nondeterminism. *Electronic Notes in Theoretical Computer Science*, 96:7–28, 2004.

[20] Gordon D. Plotkin and John Power. Semantics for algebraic operations. *Electr. Notes Theor. Comput. Sci.*, 45:332–345, 2001.

[21] Alexander Moshe Rabinovich. A complete axiomatisation for trace congruence of finite state behaviors. In *MFPS*, volume 802 of *Lecture Notes in Computer Science*, pages 530–543. Springer, 1993.

[22] Arto Salomaa. Two complete axiom systems for the algebra of regular events. *J. ACM*, 13(1):158–169, 1966.

[23] Alexandra Silva, Filippo Bonchi, Marcello M. Bonsangue, and Jan J. M. M. Rutten. Quantitative kleene coalgebras. *Inf. Comput.*, 209(5):822–849, 2011.

[24] Alexandra Silva and Ana Sokolova. Sound and complete axiomatization of trace semantics for probabilistic systems. *Electr. Notes Theor. Comput. Sci.*, 276:291–311, 2011.

[25] Eugene W. Stark and Scott A. Smolka. A complete axiom system for finite-state probabilistic processes. In *Proof, Language, and Interaction, Essays in Honour of Robin Milner*, MIT Press, pages 571–596, 2000.

[26] Marshall H. Stone. Postulates for the barycentric calculus. *Annali di Matematica Pura ed Applicata*, 29(1):25–30, 1949.

[27] Franck van Breugel and James Worrell. Towards Quantitative Verification of Probabilistic Transition Systems. In *ICALP*, volume 2076 of *Lecture Notes in Computer Science*, pages 421–432, 2001.

[28] Franck van Breugel and James Worrell. Approximating and computing behavioural distances in probabilistic transition systems. *Theoretical Computer Science*, 360(1-3):373–385, 2006.

A Proofs omitted in Section 5.1

We denote by $\ker(d) = \{(x, y) \mid d(x, y) = 0\}$ the *kernel* of a pseudometric d.

Proof of Lemma 5.7 We prove the two implications separately. ($\Leftarrow$) It suffices to show that the relation $R = \{(m, m') \mid \mathbf{d}_k(m, m') = 0\}$ (*i.e.*, $\ker(\mathbf{d}_k)$) is a k-probabilistic bisimulation. Clearly, R is an equivalence, and also $\ker(\Lambda(d^k))$ is so. Assume $(m, m') \in R$. By definition of Ψ^k, we have $\mathcal{K}(\Lambda_k(\mathbf{d}_k))(\tau_k(m), \tau_k(m')) = 0$. By [10, Lemma 3.1], for all $\ker(\Lambda_k(\mathbf{d}_k))$-equivalence classes $D \subseteq (\mathcal{L}^k \times M) \uplus \mathcal{L}^{<k}\mathcal{X}_\perp$, $\tau_k(m)(D) = \tau_k(m')(D)$. By definition of Λ_k, this implies that, for arbitrary $u \in \mathcal{L}^{<k}$, $\chi \in \mathcal{X}_\perp$, $\tau_k(m)(u\chi) = \tau_k(m')(u\chi)$ and, for all $w \in \mathcal{L}^k$ and $C \in M/_R$, $\tau_k(m)(\{w\} \times C) = \tau_k(m')(\{w\} \times C)$. ($\Rightarrow$) Let $R \subseteq M \times M$ be a k-probabilistic bisimulation on $\mathcal{M}$, and define $d_R \colon M \times M \to [0, 1]$ by $d_R(m, m') = 0$ if $(m, m') \in R$ and $d_R(m, m') = 1$ otherwise. We show that $\Psi^k(d_R) \sqsubseteq d_R$. If $(m, m') \notin R$, then $d_R(m, m') = 1 \geq \Psi^k(d_R)(m, m')$. If $(m, m') \in R$, then for all $w \in \mathcal{L}^k$, $u \in \mathcal{L}^{<k}$, $\chi \in \mathcal{X}_\perp$, and $C \in M/_R$, $\tau_k(m)(u\chi) = \tau_k(m')(u\chi)$, $\tau_k(m)(\{w\} \times C) = \tau_k(m')(\{w\} \times C)$. This implies that for all $\ker(\Lambda_k(d_R))$-equivalence class D, $\tau_k(m)(D) = \tau_k(m')(D)$. By [10, Lemma 3.1], $\mathcal{K}(\Lambda_k(d_R))(\tau_k(m), \tau_k(m')) = 0$. This implies that $\Psi^k(d_R) \sqsubseteq d_R$. Since $\sim_k$ is a k-bisimulation, $\Psi^k(d_{\sim_k}) \sqsubseteq d_{\sim_k}$, so that, by Tarski's fixed point theorem, $\mathbf{d}_k \sqsubseteq d_{\sim_k}$. By this and definition of $d_{\sim_k}$, $m \sim_k m'$ implies $\mathbf{d}_k(m, m') = 0$. □

Proof of Lemma 5.8 The set of real valued functions from $M \times M$ can be turned into a Banach space my means of the supremum norm $\|f\| = \sup_x |f(x)|$. For any countable decreasing chain with point-wise preorder $f_0 \sqsupseteq f_1 \sqsupseteq f_2 \sqsupseteq \dots$, such that $\bigsqcap_{i \in \omega} f_i$ exists, the topological limit $\lim_{i \to \infty} f_i$ coincides to $\bigsqcap_{i \in \omega} f_i$.

By interpreting 1-bounded pseudometrics as elements of the Banach space, to prove $\tilde{\Psi}_k$ is ω-cocontinuous, it suffices to show that $\tilde{\Psi}_k$ is monotone and continuous. Indeed, by monotonicity of $\tilde{\Psi}_k$ and what said above $\lim_{i \to \infty} d_i = \bigsqcap_{i \in \omega} d_i$ and $\lim_{i \to \infty} \tilde{\Psi}_k(d_i) = \bigsqcap_{i \in \omega} \tilde{\Psi}_k(d_i)$; by continuity $\lim_{i \to \infty} \tilde{\Psi}_k(d_i) = \tilde{\Psi}_k(\lim_{i \to \infty} d_i)$, so that $\bigsqcap_{i \in \omega} \tilde{\Psi}_k(d_i) = \tilde{\Psi}_k(\bigsqcap_{i \in \omega} d_i)$. Monotonicity of $\tilde{\Psi}_k$ follows trivially by monotonicity of $\mathcal{K}$ and Λ_k. We show continuity by proving that $\tilde{\Psi}_k$ is non-expansive, *i.e.*, for all $d, d' \colon M \times M \to [0, 1]$, $\|\tilde{\Psi}_k(d') - \tilde{\Psi}_k(d)\| \leq \|d' - d\|$. It suffices to prove that for all $d \sqsubseteq d'$ and $m, m' \in M$, $\tilde{\Psi}_k(d')(m, m') - \tilde{\Psi}_k(d)(m, m') \leq \|d' - d\|$. The case $m \sim_k m'$ holds trivially. Assume that $m \not\sim_k m'$, then

$$
\begin{aligned}
&\tilde{\Psi}_k(d')(m, m') - \tilde{\Psi}_k(d)(m, m') \\
&= \mathcal{K}(\Lambda_k(d'))(\tau_k(m), \tau_k(m')) - \mathcal{K}(\Lambda_k(d))(\tau_k(m), \tau_k(m')) && \text{(by def. } \tilde{\Psi}_k\text{)}
\end{aligned}
$$

by choosing $\omega \in \Omega(\tau_k(m), \tau_k(m'))$ such that $\mathcal{K}(\Lambda_k(d))(\tau_k(m), \tau_k(m')) = \int \Lambda_k(d) \, \mathrm{d}\omega$,

$$
\begin{aligned}
&= \mathcal{K}(\Lambda_k(d'))(\tau_k(m), \tau_k(m')) - \textstyle\int \Lambda_k(d) \, \mathrm{d}\omega \\
&\leq \textstyle\int \Lambda_k(d') \, \mathrm{d}\omega - \int \Lambda_k(d) \, \mathrm{d}\omega && \text{(by def. of } \mathcal{K}(\Lambda_k(d'))\text{)} \\
&= \textstyle\int (\Lambda_k(d') - \Lambda_k(d)) \, \mathrm{d}\omega && \text{(by linearity)}
\end{aligned}
$$

since, for all $(\alpha, \beta) \notin E = \{((w,n),(w,n')) \mid w \in \mathcal{L}^k, n, n' \in M\}$, by definition of Λ_k, we have $\Lambda_k(d')(\alpha,\beta) = \Lambda_k(d)(\alpha,\beta)$,

$$
\begin{aligned}
&= \textstyle\int_E (\Lambda_k(d') - \Lambda_k(d))\, \mathrm{d}\omega \\
&\leq \textstyle\int_E \|d' - d\|\, \mathrm{d}\omega && \text{(by def. } \Lambda_k) \\
&\leq \|d' - d\| \,. && \text{(by linearity and } \textstyle\int_E 1\, \mathrm{d}\omega \leq 1)
\end{aligned}
$$

It remains to show $\mathbf{d}_k = \bigsqcap_{i \in \omega} \tilde{\Psi}_k^i(\mathbf{1})$. By ω-cocontinuity $\bigsqcap_{i \in \omega} \tilde{\Psi}_k^i(\mathbf{1})$ is a fixed point. By Lemma 5.7 and $\mathbf{d}_k = \Psi_k(\mathbf{d}_k)$, also $\mathbf{d}_k$ is a fixed point of $\tilde{\Psi}_k$. We show that they coincide by proving that $\tilde{\Psi}_k$ has a unique fixed point.

Assume by contradiction that $\tilde{\Psi}_k$ has two fixed points d and d' such that $d \sqsubset d'$. Define $R \subseteq M \times M$ as $m \mathrel{R} m'$ iff $d'(m,m') - d(m,m') = \|d' - d\|$. By the assumption made on d and d' we have that $\|d' - d\| > 0$ and $R \cap \sim_k = \emptyset$. Consider arbitrary $m, m' \in M$ such that $m \mathrel{R} m'$, then

$$
\begin{aligned}
\|d' - d\| &= d'(m,m') - d(m,m') \\
&= \tilde{\Psi}_k(d')(m,m') - \tilde{\Psi}_k(d)(m,m') && \text{(by } d = \tilde{\Psi}_k(d) \text{ and } d' = \tilde{\Psi}_k(d')) \\
&= \Psi_k(d')(m,m') - \Psi_k(d)(m,m') && \text{(by } m \not\sim_k m' \text{ and def. } \tilde{\Psi}_k) \\
&\leq \textstyle\int_E (\Lambda_k(d') - \Lambda_k(d))\, \mathrm{d}\omega \,, && \text{(as proved above)}
\end{aligned}
$$

Observe that $(\Lambda_k(d') - \Lambda_k(d))((w,n),(w,n')) = d'(n) - d(n') \leq \|d' - d\|$, for all $n, n' \in M$ and $w \in \mathcal{L}^k$. Since $\|d' - d\| > 0$ the inequalities $\|d' - d\| \leq \int_E (\Lambda_k(d') - \Lambda_k(d))\, \mathrm{d}\omega \leq \|d' - d\|$ hold only if the support of ω is included in $E_R = \{((w,n),(w,n')) \mid w \in \mathcal{L}^k \text{ and } n \mathrel{R} n'\}$. Since the argument holds for arbitrary $m, m' \in M$ such that $m \mathrel{R} m'$, we have that R is a bisimulation, which is in contradiction with the initial assumptions. □

Proof of Theorem 5.11 The following inequalities hold

$$
\begin{aligned}
\mathbf{tv}^{\mathsf{omc}}([\![t]\!],[\![s]\!]) &= \mathbf{tv}([\![t]\!],[\![s]\!]) && \text{(def. } \mathbf{tv}^{\mathsf{omc}}) \\
&\leq \mathbf{tv}([\![t]\!],(\mathbb{U},t)) + \mathbf{tv}((\mathbb{U},t),(\mathbb{U},s)) + \mathbf{tv}((\mathbb{U},s),[\![s]\!]) && \text{(triang. ineq.)} \\
&= \mathbf{tv}((\mathbb{U},t),(\mathbb{U},s)) && \text{(Theorem 4.5 \& } \sim \subseteq \approx) \\
&= \mathbf{tv}^{\mathbb{U}}(t,s) \,. && \text{(def. } \mathbf{tv}^{\mathbb{U}})
\end{aligned}
$$

By a similar argument we also have $\mathbf{tv}^{\mathsf{omc}}([\![t]\!],[\![s]\!]) \geq \mathbf{tv}^{\mathbb{U}}(t,s)$, hence the thesis. □

B Proofs omitted in Section 5.5

Barycentric algebras are *entropic* [13] in the sense that all operations $+_e$ are affine maps, that is, for all $e, d \in [0,1]$ we have the entropic identity

$$\vdash (t +_e s) +_d (t' +_e s') \equiv_0 (t +_d t') +_e (s +_d s') \,. \tag{Entr}$$

If $t = s$, by (B2) the above reduces to the distributivity law

$$\vdash u +_d (t' +_e s') \equiv_0 (u +_d t') +_e (u +_d s') \,. \tag{DC}$$

Note that for $\varepsilon = \varepsilon'$, (IB) reduces to non-expansiveness for the operator $+_e$:

$$\{t \equiv_\varepsilon s, t' \equiv_\varepsilon s'\} \vdash t +_e t' \equiv_\varepsilon s +_e s' ,$$

and non-expansiveness always entails congruence for the operator: for $\varepsilon = \varepsilon' = 0$ in (Pref) and (IB) we obtain

$$\{t \equiv_0 s\} \vdash a.t \equiv_0 a.s , \qquad \text{(Pref-0)}$$

$$\{t \equiv_0 s, t' \equiv_0 s'\} \vdash t +_e t' \equiv_0 s +_e s' , \qquad \text{(IB-0)}$$

Proof of Theorem 5.14 By [25, Theorem 5.9] (which follows by Theorem 5.13), there exist terms $t_1, \dots, t_n$ with free names in $\overline{Y}$, such that $\vdash t \equiv_0 t_1$ and

$$\vdash t_i \equiv_0 \sum_{j=1}^{h(1,i)} p_{1,ij} \cdot s_{1,ij} , \qquad \text{for all } i \leq n , \qquad \text{(B.1)}$$

where the terms $s_{1,ij}$ are enumerated without repetitions, and $s_{1,ij}$ is either of the form $a_{ij}.t_{f(1,i,j)}$, or $\mathsf{rec}\, X.X$, or $Y_{g(1,i,j)}$.

By induction on $k \geq 1$, we prove that there exist terms $s_{k,ij}$ either of the form $a^1_{ij}.\dots.a^k_{ij}.t_{f(k,i,j)}$, or $b^1_{ij}.\dots.b^{u_{ij}}_{ij}.\,\mathsf{rec}\, X.X$, or $c^1_{ij}.\dots.c^{w_{ij}}_{ij}.Y_{g(k,i,j)}$, for $u_{ij}, w_{ij} < k$, s.t.

$$\vdash t_i \equiv_0 \sum_{j=1}^{h(k,i)} p_{k,ij} \cdot s_{k,ij} , \qquad \text{for all } i \leq n , \qquad \text{(B.2)}$$

where $s_{k,ij}$ are enumerated without repetitions in the formal sum above. (Base case $k = 1$) It is (B.1). (Inductive step $k > 1$) By inductive hypothesis (B.2) holds. By (Pref-0) and (DP), for each $s_{1,ij}$ of the form $a_{ij}.t_{f(1,i,j)}$ in (B.1) we have

$$\vdash s_{1,ij} \equiv_0 \sum_{e=1}^{h(k,f(1,i,j))} p_{k,f(1,i,j),e} \cdot \big(a_{ij}.s_{k,f(1,i,j),e}\big) . \qquad \text{(B.3)}$$

By (IB-0), we can substitute each $s_{1,ij}$ of the form $a_{ij}.t_{f(1,i,j)}$ in (B.1), with the right hand side of the corresponding equation (B.3). All repetitions of terms that have been introduced after the substitution can be removed by (B2), (SC), (SA). This results in the following deductions

$$\vdash t_i \equiv_0 \sum_{j=1}^{h(k+1,i)} p'_{k+1,ij} \cdot s'_{k+1,ij} , \qquad \text{for all } i \leq n , \qquad \text{(B.4)}$$

for some $h(k+1, i) \geq h(k, i)$ and $s'_{k+1,ij}$ either of the form $a^1_{ij}.\dots.a^{k+1}_{ij}.t_{f(k+1,i,j)}$, or $b^1_{ij}.\dots.b^{u_{ij}}_{ij}.\,\mathsf{rec}\, X.X$, or $c^1_{ij}.\dots.c^{w_{ij}}_{ij}.Y_{g(k+1,i,j)}$, for $u_{ij}, w_{ij} < k + 1$. □

Proof of Lemma 5.15 We proceed by well-founded induction on the strict preorder $(\mu, \nu) \prec (\mu', \nu')$ iff either $supp(\mu) \subset supp(\mu')$ or $supp(\nu) \subset supp(\nu')$.

(Base case: $supp(\mu) = \{t_1\}$ and $supp(\mu') = \{s_1\}$). In this case $\mu = \mathbb{1}_{\{t_1\}}$ and $\nu = \mathbb{1}_{\{s_1\}}$. Thus the proof follows by (Assum), by noticing that $\mathcal{K}(d)(\mu, \nu) = d(t_1, s_1)$.

(Inductive step: $supp(\mu) = \{t_1, \dots, t_k\}$ and $supp(\nu) = \{s_1, \dots, s_r\}$) Assume without loss of generality that $k > 1$ (if $k = 1$, then $r > 1$ and we proceed dually). The proof is structured as follows. We find suitable $e \in (0, 1)$ and $\mu_1, \mu_2, \nu_1, \nu_2 \in \Delta(\mathbb{T})$, such that

(i) $(\mu_1, \nu_1) \prec (\mu, \nu)$ and $(\mu_2, \nu_2) \prec (\mu, \nu)$;

(ii) $\mathcal{K}(d)(\mu, \nu) = e\mathcal{K}(d)(\mu_1, \nu_1) + (1 - e)\mathcal{K}(d)(\mu_2, \nu_2)$;

(iii) and the following are deducible

$$\vdash \textstyle\sum_{i=1}^{k} \mu(t_i) \cdot t_i \equiv_0 \left(\sum_{i=1}^{k} \mu_1(t_i) \cdot t_i \right) +_e \left(\sum_{i=1}^{k} \mu_2(t_i) \cdot t_i \right), \tag{B.5}$$

$$\vdash \textstyle\sum_{u=1}^{r} \nu(s_u) \cdot s_u \equiv_0 \left(\sum_{u=1}^{r} \nu_1(s_u) \cdot s_u \right) +_e \left(\sum_{u=1}^{r} \nu_2(s_u) \cdot s_u \right). \tag{B.6}$$

By (i) and the inductive hypothesis, we have that, for $j \in \{1, 2\}$

$$\{t_i \equiv_\varepsilon s_u \mid \varepsilon \geq d(t_i, s_u)\} \vdash \sum_{i=1}^{k} \mu_j(t_i) \cdot t_i \equiv_{\varepsilon'} \sum_{u=1}^{r} \nu_j(s_u) \cdot s_u\,, \quad \text{for all } \varepsilon' \geq \mathcal{K}(d)(\mu_j, \nu_j)\,.$$

From the above, (iii), and (IB) we deduce

$$\{t_i \equiv_\varepsilon s_u \mid \varepsilon \geq d(t_i, s_u)\} \vdash \sum_{i=1}^{k} \mu(t_i) \cdot t_i \equiv_{\varepsilon'} \sum_{u=1}^{r} \nu(s_u) \cdot s_u\,, \qquad \text{for all } \varepsilon' \geq \kappa\,.$$

where $\kappa = e\mathcal{K}(d)(\mu_1, \nu_1) + (1 - e)\mathcal{K}(d)(\mu_2, \nu_2)$. Then, the proof follows from (ii).

In the following we provide the definitions for $e \in (0, 1)$ and $\mu_1, \mu_2, \nu_1, \nu_2 \in \Delta(\mathbb{T})$, then in turn we prove (i), (ii), and (iii). Let $e = \mu(t_1)$. Note that since $supp(\mu) = \{t_1, \dots, t_k\}$ and $k > 1$, we have that $\mu(t_1) \in (0, 1)$. Let $\tilde{\omega} \in \Omega(\mu, \nu)$ be the minimal coupling for $\mathcal{K}(d)(\mu, \nu)$, *i.e.*, the one realizing the equality

$$\mathcal{K}(d)(\mu, \nu) = \textstyle\sum_{i \leq k,\, u \leq r} d(t_i, s_u) \cdot \tilde{\omega}(t_i, s_u)\,, \tag{B.7}$$

and, for $2 \leq i \leq k$, $1 \leq u \leq r$, define

$$\mu_1(t_1) = 1\,, \quad \mu_2(t_i) = \frac{\mu(t_i)}{\mu(t_1)}\,, \quad \nu_1(s_u) = \frac{\tilde{\omega}(t_1, s_u)}{\mu(t_1)}\,, \quad \nu_2(s_u) = \frac{\nu(s_u) - \tilde{\omega}(t_1, s_u)}{1 - \mu(t_1)}\,.$$

Note that $supp(\mu_1) = \{t_1\}$, $supp(\mu_2) = \{t_2, \dots, t_k\}$ and $supp(\nu_1) = supp(\nu_2) = supp(\nu)$. It is easy to show that, since $\tilde{\omega} \in \Omega(\mu, \nu)$, the above are well-defined probability distributions.

(i) It follows directly by definition of $\prec$.

(ii) For $2 \leq i \leq k$ and $1 \leq u \leq r$, define

$$\tilde{\omega}_1(t_1, s_u) = \frac{\tilde{\omega}(t_1, s_u)}{\mu(t_1)} \qquad\qquad \tilde{\omega}_2(t_i, s_u) = \frac{\tilde{\omega}(t_i, s_u)}{1 - \mu(t_1)}.$$

By a straightforward calculation, from $\tilde{\omega} \in \Omega(\mu, \nu)$ we have that $\tilde{\omega}_1 \in \Omega(\mu_1, \nu_1)$ and $\tilde{\omega}_2 \in \Omega(\mu_2, \nu_2)$. (note that this also implies that $\tilde{\omega}_1$ and $\tilde{\omega}_2$ are well-defined probability distributions on $\mathbb{T}$). From this we get the following inequality:

$$\begin{aligned}\mathcal{K}(d)(\mu_1,\nu_1) &= \sum_{i,u} d(t_i,s_u)\cdot\tilde{\omega}(t_i,s_u) && \text{(by Equation B.7)}\\ &= \mu(t_1)\left(\sum_{i,u} d(t_i,s_u)\cdot\tilde{\omega}_1(t_i,s_u)\right) + (1-\mu(t_1))\left(\sum_{i,u} d(t_i,s_u)\cdot\tilde{\omega}_2(t_i,s_u)\right)\\ &\geq \mu(t_1)\mathcal{K}(d)(\mu_1,\nu_1) + (1-\mu(t_1))\mathcal{K}(d)(\mu_2,\nu_2)\end{aligned}$$

By the above we deduce that $\tilde{\omega}_j$ is the minimal coupling for $\mathcal{K}(d)(\mu_j,\nu_j)$, for each $j\in\{1,2\}$. Indeed, if they were not minimal, the last inequality could be replaced by $<$, thus producing a contradiction. Therefore, (ii) holds.

(iii) We start by showing (B.5). Since $\mu(t_1)\in(0,1)$, the formal sum on the left-hand side of (B.5) is syntactically equivalent to

$$\sum_{i=1}^{k}\mu(t_i)\cdot t_i = t_1 +_{\mu(t_1)}\left(\sum_{i=2}^{k}\frac{\mu(t_i)}{\mu(t_1)}\cdot t_i\right), \tag{B.8}$$

By (B1), (SC), and the definitions of μ_1,μ_2 we easily obtain

$$\vdash t_1 \equiv_0 \sum_{i=1}^{k}\mu_1(t_i)\cdot t_i\,, \qquad\qquad \vdash \sum_{i=2}^{k}\frac{\mu(t_i)}{\mu(t_1)}\cdot t_i \equiv_0 \sum_{i=1}^{k}\mu_2(t_i)\cdot t_i\,.$$

Thus (B.5) follows from the deductions above by applying (IB-0) to (B.8). Next we prove (B.6) by showing that for any coupling $\omega\in\Omega(\mu,\nu)$ the following is deducible:

$$\vdash \sum_{u=1}^{r}\nu(s_u)\cdot s_u \equiv_0 \left(\sum_{u=1}^{r}\frac{\omega(t_1,s_u)}{\mu(t_1)}\cdot s_u\right) +_{\mu(t_1)} \left(\sum_{u=1}^{r}\frac{\nu(s_u)-\omega(t_1,s_u)}{1-\mu(t_1)}\cdot s_u\right). \tag{B.9}$$

We do this by induction on the size of the support of ν. (Base case: $supp(\nu)=\{s_1\}$). Then, $\nu(s_1)=1$ and $\omega(t_1,s_1)=\mu(t_1)$, so (B.9) reduces to (B2). (Inductive step: $r>1$ and $supp(\nu)=\{s_1,\dots,s_r\}$). Then $\nu(s_1)\in(0,1)$. Thus, the formal sum on the left-hand side of (B.9) is syntactically equivalent to

$$\sum_{u=1}^{r}\nu(s_u)\cdot s_u = s_1 +_{\nu(s_1)}\left(\sum_{u=2}^{r}\nu'(s_u)\cdot s_u\right), \tag{B.10}$$

where $\nu'(s_u)=\frac{\nu(s_u)}{1-\nu(s_1)}$, for $2\leq u\leq r$. Note that $\omega'(t_i,s_u)=\frac{\omega(t_i,s_u)}{1-\nu(s_1)}$, for $1\leq i\leq k$ and $2\leq u\leq r$, is a coupling in $\Omega(\mu,\nu')$ and that $supp(\nu')=\{s_2,\dots,s_r\}$. Thus, by inductive hypothesis on ν' we obtain

$$\begin{aligned}\vdash \sum_{u=2}^{r}\nu'(s_u)\cdot s_u &\equiv_0 \left(\sum_{u=2}^{r}\frac{\omega'(t_1,s_u)}{\mu(t_1)}\cdot s_u\right) +_{\mu(t_1)} \left(\sum_{u=2}^{r}\frac{\nu'(s_u)-\omega'(t_1,s_u)}{1-\mu(t_1)}\cdot s_u\right)\\ &= \left(\sum_{u=2}^{r}\frac{\omega(t_1,s_u)}{\mu(t_1)(1-\nu(s_1))}\cdot s_u\right) +_{\mu(t_1)} \left(\sum_{u=2}^{r}\frac{\nu(s_u)-\omega(t_1,s_u)}{(1-\mu(t_1))(1-\nu(s_1))}\cdot s_u\right)\end{aligned}$$

From this deduction and (B.10), by (DC), we obtain (B.9). □

Proof of Theorem 6.2 Before proving (i) and (ii), notice that, by definition of $d_{\mathbb{T}}$ and Theorems 5.12, 5.16, for any $t, s \in \mathbb{T}$, $d_{\mathbb{T}}(t,s) = \mathbf{tv}([\![t]\!], [\![s]\!])$. (i) Given $(\mathcal{M}, m), (\mathcal{N}, n)$ finite pointed open Markov chains, we construct the Σ-terms t, s as in [4, Corollary 14], obtaining that $[\![t]\!] \sim (\mathcal{M}, m)$, $[\![s]\!] \sim (\mathcal{N}, n)$. Now the thesis follows by $d_{\mathbb{T}}(t,s) = \mathbf{tv}([\![t]\!], [\![s]\!])$ and Theorem 5.11. (ii) Given $t, s \in \mathbb{T}$, we construct the finite pointed Markov chains $(\mathcal{M}, m), (\mathcal{N}, n)$ as in [4, Corollary 15], obtaining that $[\![t]\!] \sim (\mathcal{M}, m)$, $[\![s]\!] \sim (\mathcal{N}, n)$. Again the thesis follows by $d_{\mathbb{T}}(t,s) = \mathbf{tv}([\![t]\!], [\![s]\!])$ and Theorem 5.11. □